RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.

Application Serial Number: _10/549,603_
Source: _/FWP_
Date Processed by STIC: _12/1/06_

ENTERED

RAW SEQUENCE LISTING   DATE: 12/01/2006
PATENT APPLICATION: **US/10/549,603**   TIME: 14:19:55

Input Set : **A:\797_2_PCT_SeqList.TXT**
Output Set: **N:\CRF4\12012006\J549603.raw**

```
  4 <110> APPLICANT: Genencor International, Inc.
  5        Jones, Brian E.
  6        Grant, William D.
  7        Heaphy, Shaun
  8        Grant, Susan
  9        Rees, Helen
 11 <120> TITLE OF INVENTION: Novel Bacillus mHKcel Cellulase
 14 <130> FILE REFERENCE: GC797-2-PCT
C--> 16 <140> CURRENT APPLICATION NUMBER: US/10/549,603
C--> 17 <141> CURRENT FILING DATE: 2005-09-20
 19 <150> PRIOR APPLICATION NUMBER: US 60/467,315
 20 <151> PRIOR FILING DATE: 2003-04-30
 22 <160> NUMBER OF SEQ ID NOS: 3
 24 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 26 <210> SEQ ID NO: 1
 27 <211> LENGTH: 3796
 28 <212> TYPE: DNA
 29 <213> ORGANISM: Bacillus sp.
 31 <220> FEATURE:
 32 <221> NAME/KEY: misc_feature
 33 <222> LOCATION: (1)...(3796)
 34 <223> OTHER INFORMATION: isolated from environmental water sample from
 35        Sonachi Lake, Kenya
 37 <221> NAME/KEY: misc_feature
 38 <222> LOCATION: (1)...(3796)
 39 <223> OTHER INFORMATION: n = A,T,C or G
W--> 41 <400> 1
 42 atctttagag atgagcgtga tagtcagtgg accttaaaag aacttaaaag atttttagaa        60
 43 gagattaaaa cggacccgca ccatctctct gtttattttg atgggggatt tgatttggag       120
 44 acacaacgat ctggccttgg gtgtgtgatt tattatgaac aaaatgacac gtcttaccgg       180
 45 gtgagaaaaa acgctaccgt ggcgtcattg acatcgaata accaagcaga atatgccgct       240
 46 ttgcatttag gacttaaaag aacttgaagg gatcgggcgc atcatctgcc tatcaccatt       300
 47 tacggtgatt ctcaagttgt gatcaatcag ttaaaaggag aatgggccgt gtatggagga       360
 48 ggtgttaaat aaatgggctg acccgtattg atcaagcatt tagctaaatt aggcatgacc       420
 49 gctacttata agttaatccc ccgtaaagaa aaccgcgaag ccgatcaact ggctacacaa       480
 50 gcgttaaacg ggcaagaaat tataagtcaa cgtgatatca gtgagcgtgg tgcagattag       540
 51 gctgcaccgc gcaaaaaaag tcaacgtgtt taggaatgga cagggattaa agcaacataa       600
 52 ttctctctaa gcaaacgttg cgacagcaag agagaagcat ataaggtttt tctgagttag       660
 53 tctatttata ccaatgtcca cgtactaaat aaacctctca tcaaagtgga tttttttgatt      720
 54 aattcacttc cactcctacc tttatctata taaattagtt ccttttttgt taataatcac       780
 55 taattttggc ggtattttt aatagaaata tatgctagat tataaactag tagccgtata        840
 56 gaaggtggtg attgcccta taagagacgt ctggcaaaca taaaagcatc gcattattat        900
 57 aatcgaaagg tggagatgag acatgggtta tacccaagct aagtgtatgg tgaaaaaaac       960
```

RAW SEQUENCE LISTING   DATE: 12/01/2006
PATENT APPLICATION: US/10/549,603   TIME: 14:19:55

Input Set : A:\797_2_PCT_SeqList.TXT
Output Set: N:\CRF4\12012006\J549603.raw

```
     58 ggtcttgttt ggtttaattc tctgtttagg tgtgtcaatg tttgtaccag ttacatcagc      1020
     59 tgaagatagg gtctcttcgt cacaggtgga tatccaatca tatgtagcag atatgcaacc      1080
     60 tggctggaat ttaggtaata catttgatgc gataggagat gatgaaacag catgggaaa      1140
     61 ccctcgtgta acgagagaat taatagaaat gattgctgat gaagggtata aaagtattcg      1200
     62 tatcccagtc acatggcaaa atcaaatggg tggttctcca gattatacaa ttaatgaaga      1260
     63 ttatatcaag cgggtagagc aagtgataga ttgggcgttg gaggaagact tgtatgtgat      1320
     64 gttaaatgtg catcatgact catggctgtg gatgtatgat atggaacata actatgatga      1380
     65 ggtgatggca agatatacag ctatttggga acaattgtcg gaaaaattca aaaaccactc      1440
     66 ccataagttg atgtttgaga gtgtcaatga gcctaggttt acgcaggagt ggggagagat      1500
     67 tcaagaaaat catcatgctt acttagaaga tttaaataag acgttctatt atattgtcag      1560
     68 agagtcagga ggcaataatg tggagcgccc tttagtattg cctacgatag aaacagccac      1620
     69 gtctcaggat ttactagatc gcttgtatca aacaatggaa gacttggatg accctcattt      1680
     70 aattgccacg gttcattatt atggcttttg gccctttagt gtcaatatag cagggtacac      1740
     71 ccgttttgaa caggagacac aacaagatat tatagacacg tttgaccgtg ttcataacac      1800
     72 atttacagcg aatgggatcc cagttgtatt aggtgaattt ggtttgttag gctttgataa      1860
     73 aagtacggac gtcattcagc aaggtgagaa attaaaattt tttgagtttc tcatccatca      1920
     74 tctcaatgaa cgtgatataa cccatatgtt atgggataac ggtcagcatt taaagcgaga      1980
     75 aacttattca tggtatgatc aggaatttca tgacatatta aaagcgagtt gggaggggcg      2040
     76 ttctgctaca gctgagtcta atttcattca tgtgaaggac ggagagccaa ttagagatca      2100
     77 acatatacag ctttacttaa acggaaatga gctaactgcc ctacaggcag gggacgaatc      2160
     78 gcttgtacta ggagaggatt atgagctagc aggagacgta ttaacgctaa aagcgggcat      2220
     79 cctcacaaga ttaattaccc ctggccaatt aggaacgaat gcggtcatca cagctcaatt      2280
     80 taattctgga gcagactggc gttttcaatt acagaatgtg gacgtgccaa cagtcgaaaa      2340
     81 tacagatggc tcaatatggc attttgcgat ccctacccat tttaatggtg atagtcttgc      2400
     82 gacgatggaa gctgtttatg caaacggaga atatgctggc ccgcaagatt ggacgtcatt      2460
     83 taaagaattt ggcgaggcgt tttcccctaa ttacgccaca ggggaaatta ttataacaga      2520
     84 agccttcttt aacgcggtac gggatgatga tatccattta acatttcatt attggagcgg      2580
     85 agagacggtg gaatatacat tacgtaaaaa tggaaattat gttcaaggta gacggtaaca      2640
     86 tgattttaat taatagataa aacagcctac ctatcgtttt tggaagaagg caaacgaatc      2700
     87 tcatcttacc aacaccgtgc tttagaactt tagaagtgac ggtgtttttt aagacatgag      2760
     88 gagagacaat cctctatcaa cagtcaccaa tttttattca ggaggtgtca agttatctaa      2820
     89 cgttctatga atgcatatag tttctgacga ataaacatag ttaaaaagaa gtgagcctag      2880
     90 ttcccgaggg gaaggggata atgccaacgt attggattaa agtaccttct tgataaaaag      2940
     91 aaagggtttt caagaggtgg aaatgggctc gtttgttata cttttaattac accttggaac      3000
     92 gtcattttgg cggtgctact tagtaagatg actgacatca taaaagagga gtgggttcga      3060
     93 tggctttaat tcaattaagc tttaaatcac gagcattaat gttgcaaacc tctgtcaatg      3120
     94 ttttattacc ggtgggaatg aatgcggtag attttacacc aagtgatgat ttttcttatg      3180
     95 ttactgaccc ttttcctgtc ctatatcttt tgcatggtgc aactgatgat tattcagcat      3240
     96 ggctacgtct gtcctccatt gaacgatatg ctgaagaaaa aaaattggcg gtcgtcatgc      3300
     97 caaatgctga tatgagtgcg tatacggata tggtacatgg acatcgttac tggacgtata      3360
     98 ttagtaagga gctgcctgag tttatcaaag cgactttcc tatttctcag caccgtgaag      3420
     99 acacctttgc ggctggtctg tctatgggag gatacggggc ttttaaatta gcgttgcggc      3480
W--> 100 aaccggaacg cttcgctgca gntgtgtcat tatcaggtgc agttgatatg agagaagcaa      3540
    101 gtcaaccaga ctccctattt gtgaacgcat ttggtgaagg gacgaaaatc gcagggacag      3600
    102 atcttgatct tttttcattta attaaaaagt tgggggtata tgaaggggct aaaccagccc      3660
    103 tttttcaagc gtgtgggaca gaggacttt tatatgaaga taatgtgaga tttagagatt      3720
    104 atgcacgaca agtgaatgcc gatttaactt atgaagaagg tcctggtggt catgaatggg      3780
    105 cttattggga tagaat                                                      3796
    107 <210> SEQ ID NO: 2
```

RAW SEQUENCE LISTING   DATE: 12/01/2006
PATENT APPLICATION: US/10/549,603   TIME: 14:19:55

Input Set : **A:\797_2_PCT_SeqList.TXT**
Output Set: **N:\CRF4\12012006\J549603.raw**

```
108 <211> LENGTH: 1716
109 <212> TYPE: DNA
110 <213> ORGANISM: Bacillus sp.
112 <220> FEATURE:
113 <221> NAME/KEY: misc_feature
114 <222> LOCATION: (1)...(1716)
115 <223> OTHER INFORMATION: isolated from environmental water sample from
116         Sonachi Lake, Kenya
118 <400> SEQUENCE: 2
119 atgggttata cccaagctaa gtgtatggtg aaaaaaacgg tcttgtttgg tttaattctc       60
120 tgtttaggtg tgtcaatgtt tgtaccagtt acatcagctg aagatagggt ctcttcgtca      120
121 caggtggata tccaatcata tgtagcagat atgcaacctg ctggaatttt aggtaataca      180
122 tttgatgcga taggagatga tgaaacagca tggggaaacc ctcgtgtaac gagagaatta      240
123 atagaaatga ttgctgatga agggtataaa agtattcgta tcccagtcac atggcaaaat      300
124 caaatgggtg gttctccaga ttatacaatt aatgaagatt atatcaagcg ggtagagcaa      360
125 gtgatagatt gggcgttgga ggaagacttg tatgtgatgt aaatgtgca tcatgactca       420
126 tggctgtgga tgtatgatat ggaacataac tatgatgagg tgatggcaag atatacagct      480
127 atttgggaac aattgtcgga aaaattcaaa aaccactccc ataagttgat gtttgagagt      540
128 gtcaatgagc ctaggtttac gcaggagtgg ggagagattc aagaaaatca tcatgcttac      600
129 ttagaagatt taaataagac gttctattat attgtcagag agtcaggagg caataatgtg      660
130 gagcgccctt tagtattgcc tacgatagaa acagccacgt ctcaggattt actagatcgc      720
131 ttgtatcaaa caatggaaga cttggatgac cctcatttaa ttgccacggt tcattattat      780
132 ggcttttggc cctttagtgt caatatagca gggtacaccc gttttgaaca ggagacacaa      840
133 caagatatta tagacacgtt tgaccgtgtt cataacacat ttacagcgaa tgggatccca      900
134 gttgtattag gtgaatttgg tttgttaggc tttgataaaa gtacggacgt cattcagcaa      960
135 ggtgagaaat taaaattttt tgagtttctc atccatcatc tcaatgaacg tgatataacc     1020
136 catatgttat gggataacgg tcagcattta aagcgagaaa cttattcatg gtatgatcag     1080
137 gaatttcatg acatattaaa agcgagttgg gaggggcgtt ctgctacagc tgagtctaat     1140
138 ttcattcatg tgaaggacgg agagccaatt agagatcaac atatacagct ttacttaaac     1200
139 ggaaatgagc taactgccct acaggcaggg gacgaatcgc ttgtactagg agaggattat     1260
140 gagctagcag gagacgtatt aacgctaaaa gcgggcatcc tcacaagatt aattacccct     1320
141 ggccaattag gaacgaatgc ggtcatcaca gctcaattta attctggagc agactggcgt     1380
142 tttcaattac agaatgtgga cgtgccaaca gtcgaaaata cagatggctc aatatggcat     1440
143 tttgcgatcc ctacccattt taatggtgat agtcttgcga cgatggaagc tgtttatgca     1500
144 aacggagaat atgctggccc gcaagattgg acgtcattta aagaatttgg cgaggcgttt     1560
145 tcccctaatt acgccacagg ggaaattatt ataacagaag ccttctttaa cgcggtacgg     1620
146 gatgatgata tccatttaac atttcattat tggagcggag agacggtgga atatacatta     1680
147 cgtaaaaatg gaaattatgt tcaaggtaga cggtaa                               1716
149 <210> SEQ ID NO: 3
150 <211> LENGTH: 571
151 <212> TYPE: PRT
152 <213> ORGANISM: Bacillus sp.
154 <220> FEATURE:
155 <221> NAME/KEY: VARIANT
156 <222> LOCATION: (1)...(571)
157 <223> OTHER INFORMATION: isolated from environmental water sample from
158         Sonachi Lake, Kenya
160 <400> SEQUENCE: 3
161 Met Gly Tyr Thr Gln Ala Lys Cys Met Val Lys Lys Thr Val Leu Phe
```

RAW SEQUENCE LISTING   DATE: 12/01/2006
PATENT APPLICATION: US/10/549,603   TIME: 14:19:56

Input Set : A:\797_2_PCT_SeqList.TXT
Output Set: N:\CRF4\12012006\J549603.raw

```
162  1                   5                   10                  15
163 Gly Leu Ile Leu Cys Leu Gly Val Ser Met Phe Val Pro Val Thr Ser
164             20                  25                  30
165 Ala Glu Asp Arg Val Ser Ser Ser Gln Val Asp Ile Gln Ser Tyr Val
166             35                  40                  45
167 Ala Asp Met Gln Pro Gly Trp Asn Leu Gly Asn Thr Phe Asp Ala Ile
168        50                  55                  60
169 Gly Asp Asp Glu Thr Ala Trp Gly Asn Pro Arg Val Thr Arg Glu Leu
170 65                  70                  75                  80
171 Ile Glu Met Ile Ala Asp Glu Gly Tyr Lys Ser Ile Arg Ile Pro Val
172             85                  90                  95
173 Thr Trp Gln Asn Gln Met Gly Gly Ser Pro Asp Tyr Thr Ile Asn Glu
174            100                 105                 110
175 Asp Tyr Ile Lys Arg Val Glu Gln Val Ile Asp Trp Ala Leu Glu Glu
176            115                 120                 125
177 Asp Leu Tyr Val Met Leu Asn Val His His Asp Ser Trp Leu Trp Met
178        130                 135                 140
179 Tyr Asp Met Glu His Asn Tyr Asp Glu Val Met Ala Arg Tyr Thr Ala
180 145                 150                 155                 160
181 Ile Trp Glu Gln Leu Ser Glu Lys Phe Lys Asn His Ser His Lys Leu
182            165                 170                 175
183 Met Phe Glu Ser Val Asn Glu Pro Arg Phe Thr Gln Glu Trp Gly Glu
184            180                 185                 190
185 Ile Gln Glu Asn His His Ala Tyr Leu Glu Asp Leu Asn Lys Thr Phe
186            195                 200                 205
187 Tyr Tyr Ile Val Arg Glu Ser Gly Gly Asn Asn Val Glu Arg Pro Leu
188        210                 215                 220
189 Val Leu Pro Thr Ile Glu Thr Ala Thr Ser Gln Asp Leu Leu Asp Arg
190 225                 230                 235                 240
191 Leu Tyr Gln Thr Met Glu Asp Leu Asp Asp Pro His Leu Ile Ala Thr
192            245                 250                 255
193 Val His Tyr Tyr Gly Phe Trp Pro Phe Ser Val Asn Ile Ala Gly Tyr
194            260                 265                 270
195 Thr Arg Phe Glu Gln Glu Thr Gln Gln Asp Ile Ile Asp Thr Phe Asp
196        275                 280                 285
197 Arg Val His Asn Thr Phe Thr Ala Asn Gly Ile Pro Val Val Leu Gly
198        290                 295                 300
199 Glu Phe Gly Leu Leu Gly Phe Asp Lys Ser Thr Asp Val Ile Gln Gln
200 305                 310                 315                 320
201 Gly Glu Lys Leu Lys Phe Phe Glu Phe Leu Ile His His Leu Asn Glu
202            325                 330                 335
203 Arg Asp Ile Thr His Met Leu Trp Asp Asn Gly Gln His Leu Lys Arg
204            340                 345                 350
205 Glu Thr Tyr Ser Trp Tyr Asp Gln Glu Phe His Asp Ile Leu Lys Ala
206            355                 360                 365
207 Ser Trp Glu Gly Arg Ser Ala Thr Ala Glu Ser Asn Phe Ile His Val
208        370                 375                 380
209 Lys Asp Gly Glu Pro Ile Arg Asp Gln His Ile Gln Leu Tyr Leu Asn
210 385                 390                 395                 400
```

RAW SEQUENCE LISTING   DATE: 12/01/2006
PATENT APPLICATION: **US/10/549,603**   TIME: 14:19:56

Input Set : **A:\797_2_PCT_SeqList.TXT**
Output Set: **N:\CRF4\12012006\J549603.raw**

```
211 Gly Asn Glu Leu Thr Ala Leu Gln Ala Gly Asp Glu Ser Leu Val Leu
212                 405                 410                 415
213 Gly Glu Asp Tyr Glu Leu Ala Gly Asp Val Leu Thr Leu Lys Ala Gly
214                 420                 425                 430
215 Ile Leu Thr Arg Leu Ile Thr Pro Gly Gln Leu Gly Thr Asn Ala Val
216             435                 440                 445
217 Ile Thr Ala Gln Phe Asn Ser Gly Ala Asp Trp Arg Phe Gln Leu Gln
218       450                 455                 460
219 Asn Val Asp Val Pro Thr Val Glu Asn Thr Asp Gly Ser Ile Trp His
220 465                 470                 475                 480
221 Phe Ala Ile Pro Thr His Phe Asn Gly Asp Ser Leu Ala Thr Met Glu
222             485                 490                 495
223 Ala Val Tyr Ala Asn Gly Glu Tyr Ala Gly Pro Gln Asp Trp Thr Ser
224             500                 505                 510
225 Phe Lys Glu Phe Gly Glu Ala Phe Ser Pro Asn Tyr Ala Thr Gly Glu
226         515                 520                 525
227 Ile Ile Ile Thr Glu Ala Phe Phe Asn Ala Val Arg Asp Asp Asp Ile
228       530                 535                 540
229 His Leu Thr Phe His Tyr Trp Ser Gly Glu Thr Val Glu Tyr Thr Leu
230 545                 550                 555                 560
231 Arg Lys Asn Gly Asn Tyr Val Gln Gly Arg Arg
232             565                 570
```

RAW SEQUENCE LISTING ERROR SUMMARY   DATE: 12/01/2006
PATENT APPLICATION: US/10/549,603   TIME: 14:19:57

Input Set : A:\797_2_PCT_SeqList.TXT
Output Set: N:\CRF4\12012006\J549603.raw

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:1; N Pos. 3502

VERIFICATION SUMMARY　　　　　　DATE: 12/01/2006
PATENT APPLICATION:　**US/10/549,603**　　TIME: 14:19:57

Input Set : **A:\797_2_PCT_SeqList.TXT**
Output Set: **N:\CRF4\12012006\J549603.raw**

L:16 M:270 C: Current Application Number differs, Replaced Current Application Number
L:17 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:41 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:100 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1 after pos.:3480